

# How Faithful is your Synthetic Data? Sample-level Metrics for Evaluating and Auditing Generative Models

Year: 2021 | Citations: 239 | Authors: A. Alaa, B. V. Breugel, Evgeny S. Saveliev, M. Schaar

---

## Abstract

Devising domain- and model-agnostic evaluation metrics for generative models is an important and as yet unresolved problem. Most existing metrics, which were tailored solely to the image synthesis setup, exhibit a limited capacity for diagnosing the different modes of failure of generative models across broader application domains. In this paper, we introduce a 3-dimensional evaluation metric, ( $\alpha$ -Precision,  $\beta$ -Recall, Authenticity), that characterizes the fidelity, diversity and generalization performance of any generative model in a domain-agnostic fashion. Our metric unifies statistical divergence measures with precision-recall analysis, enabling sample- and distribution-level diagnoses of model fidelity and diversity. We introduce generalization as an additional, independent dimension (to the fidelity-diversity trade-off) that quantifies the extent to which a model copies training data -- a crucial performance indicator when modeling sensitive data with requirements on privacy. The three metric components correspond to (interpretable) probabilistic quantities, and are estimated via sample-level binary classification. The sample-level nature of our metric inspires a novel use case which we call model auditing, wherein we judge the quality of individual samples generated by a (black-box) model, discarding low-quality samples and hence improving the overall model performance in a post-hoc manner.